

EA2020-analytical.R

Edre MA, DrPH

2020-12-02

```
#=====
#Analytical statistics
#R Biostat Workshop IIUM
#Edre MA, DrPH
#=====

#Comparing numerical values: parametric

library(readxl)

## Warning: package 'readxl' was built under R version 3.6.3

healthstat <- read_excel("healthstatus.xlsx")
View(healthstat)

#independent sample t test
t.test(sbp ~ sex, data = healthstat)

##
## Welch Two Sample t-test
##
## data: sbp by sex
## t = 0.4262, df = 145.43, p-value = 0.6706
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.860412 4.433217
## sample estimates:
## mean in group Female mean in group Male
## 124.3286 123.5422

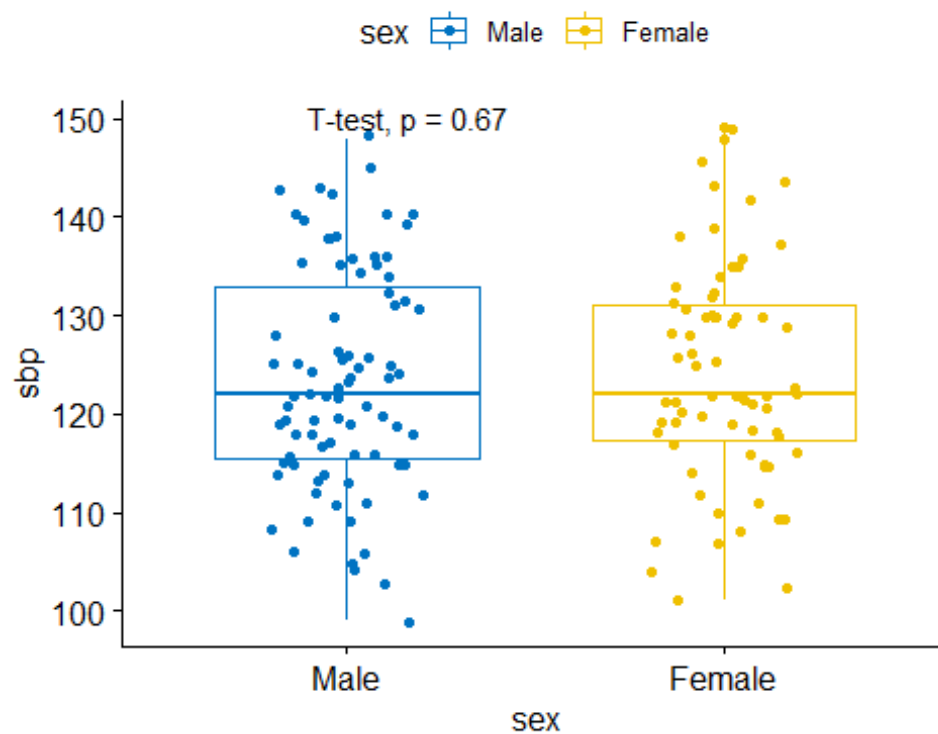
library(ggpubr)

## Warning: package 'ggpubr' was built under R version 3.6.3

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.6.3

ggboxplot(healthstat, x = "sex", y = "sbp", color = "sex", palette = "jco",
add = "jitter") + stat_compare_means(method = "t.test")
```



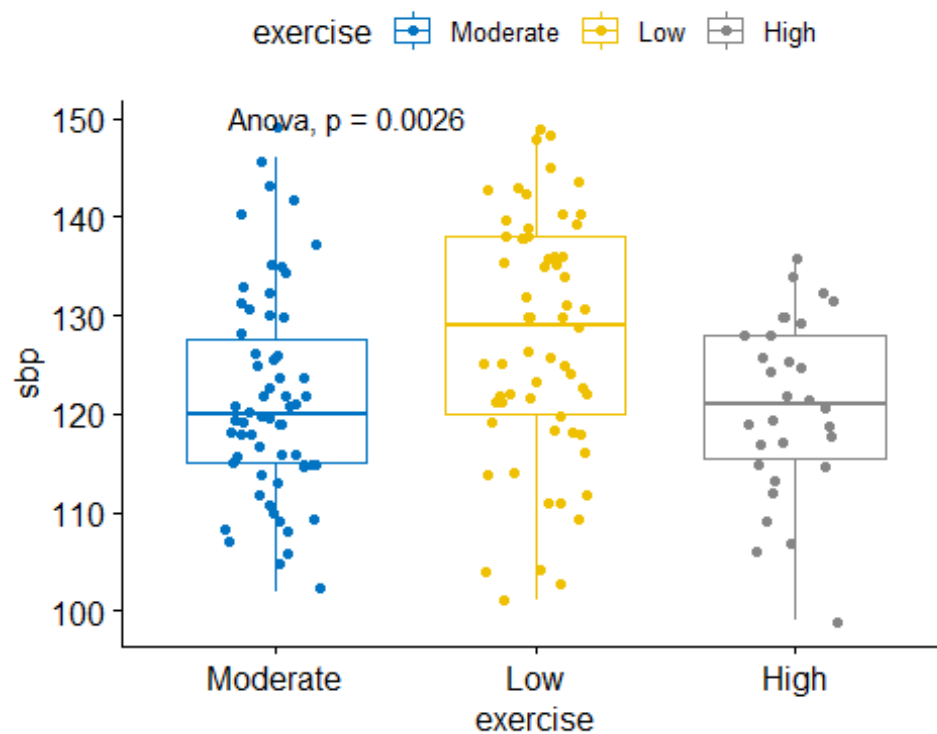
#one way ANOVA

```
one.way = aov(sbp ~ exercise, data = healthstat)
summary(one.way)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## exercise      2   1490    744.8    6.215 0.00255 **
## Residuals    150  17976    119.8
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggboxplot(healthstat, x = "exercise", y = "sbp", color = "exercise", palette =
"jco", add = "jitter") + stat_compare_means(method = "anova")
```



```
tukey.one.way<-TukeyHSD(one.way) #assuming equal variance
tukey.one.way
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = sbp ~ exercise, data = healthstat)
##
## $exercise
##          diff          lwr          upr      p adj
## Low-High    6.8213115    1.042596 12.600027 0.0161206
## Moderate-High 0.6967742 -5.066557  6.460106 0.9558683
## Moderate-Low -6.1245373 -10.797884 -1.451191 0.0064569
```

```
#paired t test
```

```
t.test(healthstat$wt, healthstat$wt2, paired=TRUE)
```

```
##
## Paired t-test
##
## data: healthstat$wt and healthstat$wt2
## t = 24.838, df = 152, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.618054 4.243384
## sample estimates:
```

```
## mean of the differences
##          3.930719

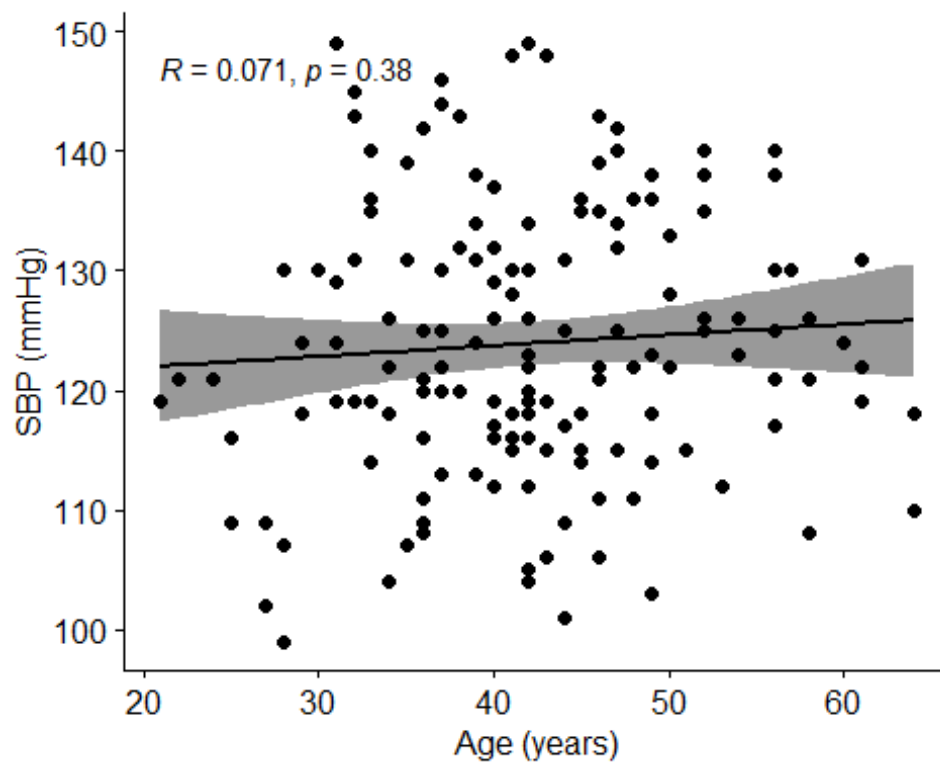
#pearson correlation coefficient test

cor.test(healthstat$age,healthstat$sbp, method="pearson")

##
## Pearson's product-moment correlation
##
## data: healthstat$age and healthstat$sbp
## t = 0.87193, df = 151, p-value = 0.3846
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.08889803  0.22690823
## sample estimates:
##          cor
## 0.07077851

ggscatter(healthstat, x = "age", y = "sbp", add = "reg.line", conf.int =
TRUE, cor.coef = TRUE, cor.method = "pearson", xlab = "Age (years)", ylab =
"SBP (mmHg)")

## `geom_smooth()` using formula 'y ~ x'
```



```
#comparing numerical values: non-parametric
```

```
#mann whitney U test
```

```
wilcox.test(hcy~sex, data=healthstat)
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: hcy by sex
```

```
## W = 2428.5, p-value = 0.08129
```

```
## alternative hypothesis: true location shift is not equal to 0
```

```
#kruskal wallis test
```

```
kruskal.test(hcy ~ exercise, data = healthstat) #if significant, proceed with pairwise comparison
```

```
##
```

```
## Kruskal-Wallis rank sum test
```

```
##
```

```
## data: hcy by exercise
```

```
## Kruskal-Wallis chi-squared = 14.376, df = 2, p-value = 0.0007555
```

```
pairwise.wilcox.test(healthstat$hcy, healthstat$exercise,p.adjust.method = "BH")
```

```
##
```

```
## Pairwise comparisons using Wilcoxon rank sum test
```

```
##
```

```
## data: healthstat$hcy and healthstat$exercise
```

```
##
```

```
##           High    Low
```

```
## Low           0.0029 -
```

```
## Moderate 0.5267 0.0032
```

```
##
```

```
## P value adjustment method: BH
```

```
#wilcoxon signed rank test
```

```
wilcox.test(healthstat$wt,healthstat$wt2,paired=TRUE)
```

```
##
```

```
## Wilcoxon signed rank test with continuity correction
```

```
##
```

```
## data: healthstat$wt and healthstat$wt2
```

```
## V = 11769, p-value < 2.2e-16
```

```
## alternative hypothesis: true location shift is not equal to 0
```

```
#spearman correlation coefficient test
```

```
cor.test(healthstat$hba1c,healthstat$sbp, method="spearman")
```

```
## Warning in cor.test.default(healthstat$hba1c, healthstat$sbp, method =  
## "spearman"): Cannot compute exact p-value with ties
```

```
##  
## Spearman's rank correlation rho  
##  
## data: healthstat$hba1c and healthstat$sbp  
## S = 495474, p-value = 0.03573  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## 0.1699267
```

#comparing categorical values

#chi square test

```
chisq.test(healthstat$sex,healthstat$smoking,correct=F)
```

```
##  
## Pearson's Chi-squared test  
##  
## data: healthstat$sex and healthstat$smoking  
## X-squared = 23.839, df = 1, p-value = 1.047e-06
```

#fisher's exact test (used when more than 20% cells with expected count less than 5)

```
chisq.test(healthstat$exercise,healthstat$smoking)$expected
```

```
##  
## healthstat$exercise healthstat$smoking  
## No Yes  
## High 20.58824 9.411765  
## Low 41.86275 19.137255  
## Moderate 42.54902 19.450980
```

```
fisher.test(healthstat$exercise,healthstat$smoking)
```

```
##  
## Fisher's Exact Test for Count Data  
##  
## data: healthstat$exercise and healthstat$smoking  
## p-value = 0.001413  
## alternative hypothesis: two.sided
```

#reporting your findings in table form

#package needed

#"sjPlot"

#"apaTables"

```
library(sjPlot)
```

```
## Warning: package 'sjPlot' was built under R version 3.6.3

## Registered S3 methods overwritten by 'lme4':
##   method                      from
##   cooks.distance.influence.merMod car
##   influence.merMod             car
##   dfbeta.influence.merMod      car
##   dfbetas.influence.merMod    car

## Install package "strengexjacke" from GitHub
(`devtools::install_github("strengexjacke/strengexjacke")`) to load all sj-
packages at once!

library(apaTables)

## Warning: package 'apaTables' was built under R version 3.6.3

#table created in word file in your directory!
sjt.xtab(healthstat$sex, healthstat$smoking, file = "sjt_contingency.doc")
```

```
sex
smoking
Total
No
Yes
Female
```

```
62
8
70
Male
43
40
83
Total
105
48
153
```

```
 $\chi^2=22.162 \cdot df=1 \cdot \varphi=0.395 \cdot p=0.000$ 
```

```
apa.aov.table(one.way, filename="Table_anova.doc", table.number = 2)
```

```
##
##
## Table 2
##
## ANOVA results using sbp as the dependent variable
##
##
```

```
##      Predictor      SS   df      MS      F      p partial_eta2
## (Intercept) 438504.30   1 438504.30 3659.11 .000
##      exercise   1489.65   2    744.83   6.22 .003          .08
##      Error   17975.88 150    119.84
## CI_90_partial_eta2
##
##      [.02, .14]
##
## Note: Values in square brackets indicate the bounds of the 90% confidence
## interval for partial eta-squared
```